Day 4B
Bootstrap

A. Colin Cameron
Univ. of Calif. - Davis
... for
Center of Labor Economics
Norwegian School of Economics
Advanced Microeconometrics

Aug 28 - Sep 2, 2017

1. Introduction

- The bootstrap is a method for obtaining properties of statistics through resampling.
- There are many ways to bootstrap.
- There are many uses of the bootstrap.
- The most common uses of the bootstrap in econometrics are
  - to obtain standard errors of estimates.
- Occasionally use a more advanced bootstrap to potentially enable better finite sample inference
  - confidence intervals with better coverage
  - tests with true size closer to nominal size.

Summary

1. Introduction
2. Bootstrap (without asymptotic refinement)
3. Bootstrap in General
4. Bootstrap with asymptotic refinement
5. Wild Bootsrap
6. Stata commands

2. Bootstrap pairs estimate of standard error

- The most common bootstrap is the pairs bootstrap
  - views the sample as the $\{(y_1, \mathbf{x}_1), ..., (y_N, \mathbf{x}_N)\}$ as the population
  - assumes that $(y_i, \mathbf{x}_i)$ are i.i.d.
  - obtains B random samples from this population by resampling with replacement
    - e.g. in bootstrap resample 1 observation may appear once, observation 2 not at all, observation 2 times, ...
- This yields B estimates $\widehat{\boldsymbol{\theta}}_1, ..., \widehat{\boldsymbol{\theta}}_B$.
  - so estimate $\mathrm{Var}[\widehat{\boldsymbol{\theta}}]$ using the usual variance of the B estimates.
- For scalar θ we have

$$\widehat{\mathrm{V}}[\widehat{\theta}] = \tfrac{1}{B-1} \textstyle\sum_{b=1}^{B} (\widehat{\theta}_b - \overline{\widehat{\theta}})^2, \quad \text{where } \overline{\widehat{\theta}} = \tfrac{1}{B} \textstyle\sum_{b=1}^{B} \widehat{\theta}_b.$$

  - Square root of this is called a bootstrap standard error.

Regression application

- Data: Doctor visits (count) and chronic conditions. $N = 50$.

```
. * Summmarize and Poisson with robust se's
. summarize

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      docvis |         50        4.12     7.82106          0         43
         age |         50       4.162    1.160382        2.6        6.2
     chronic |         50         .28    .4535574          0          1

. poisson docvis chronic, nolog vce(robust)

Poisson regression                              Number of obs     =         50
                                                Wald chi2(1)      =       3.64
                                                Prob > chi2       =     0.0565
Log pseudolikelihood = -238.75384               Pseudo R2         =     0.0917

------------------------------------------------------------------------------
             |               Robust
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     chronic |   .9833014   .5154894     1.91   0.056    -.0270391    1.993642
       _cons |   1.031602   .3446734     2.99   0.003     .3560541    1.707149
------------------------------------------------------------------------------
```

Bootstrap standard errors after Poisson regression

- Use option vce(boot)
  - Set the seed!
  - Set the number of bootstrap repetitions!

```
. * Compute bootstrap standard errors using option vce(bootstrap) to
. poisson docvis chronic, vce(boot, reps(400) seed(10101) nodots)

Poisson regression                              Number of obs     =         50
                                                Replications      =        400
                                                Wald chi2(1)      =       3.33
                                                Prob > chi2       =     0.0679
Log likelihood = -238.75384                     Pseudo R2         =     0.0917

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
      docvis |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     chronic |   .9833014   .5386575     1.83   0.068    -.0724478    2.039051
       _cons |   1.031602   .3536507     2.92   0.004      .338459    1.724744
------------------------------------------------------------------------------
```

- Bootstrap se $= 0.539$ versus White robust se $= 0.515$.

Results vary with seed and number of reps

```
. * Bootstrap standard errors for different reps and seeds
. quietly poisson docvis chronic, vce(boot, reps(50) seed(10101))

. estimates store boot50

. quietly poisson docvis chronic, vce(boot, reps(50) seed(20202))

. estimates store boot50diff

. quietly poisson docvis chronic, vce(boot, reps(2000) seed(10101))

. estimates store boot2000

. quietly poisson docvis chronic, vce(robust)

. estimates store robust

. estimates table boot50 boot50diff boot2000 robust, b(%8.5f) se(%8.5f)
```

Variable	boot50	boot50~f	boot2000	robust
chronic	0.98330	0.98330	0.98330	0.98330
	0.45444	0.59923	0.54178	0.51549
_cons	1.03160	1.03160	1.03160	1.03160
	0.37131	0.37533	0.36414	0.34467

legend: b/se

Leading uses of bootstrap standard errors

- Sequential two-step m-estimator
  - First step gives $\widehat{\alpha}$ used to create a regressor $z(\widehat{\alpha})$
  - Second step regresses y on x and $z(\widehat{\alpha})$
  - Do a paired bootstrap resampling (x, y, z)
  - e.g. Heckman two-step estimator.
- 2SLS estimator with heteroskedastic errors (if no White option)
  - Paired bootstrap gives heteroskedastic robust standard errors.
- Functions of other estimates e.g. $\widehat{\theta} = \widehat{\alpha} \times \widehat{\beta}$
  - replaces delta method
  - Clustered data with many small clusters, such as short panels.
    - Then resample the clusters.
    - But be careful if model includes cluster-specific fixed effects.
- For these in Stata need to use prefix command `bootstrap:`

3. The Bootstrap in General: Bootstrap algorithm

- A general bootstrap algorithm is as follows:
  - **1.** Given data $\mathbf{w}_1, ..., \mathbf{w}_N$
    - ★ draw a bootstrap sample of size N (see below)
    - ★ denote this new sample $\mathbf{w}_1^*, ..., \mathbf{w}_N^*$.
  - **2.** Calculate an appropriate statistic using the bootstrap sample. Examples include:
    - ★ (a) estimate $\widehat{\theta}^*$ of θ;
    - ★ (b) standard error $s_{\widehat{\theta}^*}$ of estimate $\widehat{\theta}^*$
    - ★ (c) $t-$statistic $t^* = (\widehat{\theta}^* - \widehat{\theta})/s_{\widehat{\theta}^*}$ centered at $\widehat{\theta}$.
  - **3.** Repeat steps 1-2 B independent times.
    - ★ Gives B bootstrap replications of $\widehat{\theta}_1^*, ..., \widehat{\theta}_B^*$ or $t_1^*, \ldots, t_B^*$ or .....
  - **4.** Use these B bootstrap replications to obtain a bootstrapped version of the statistic (see below).

Implementation

- Number of bootstraps: B high is best but increases computer time.
  - CT use 400 for se's and 999 for tests and confidence intervals.
  - Defaults are often too low. And set the seed!
- Various resampling methods
  - 1. Paired (or nonparametric or empirical dist. func.) is most common
    - $\mathbf{w}_1^*, ..., \mathbf{w}_N^*$ obtained by sampling with replacement from $\mathbf{w}_1, ..., \mathbf{w}_N$.
  - 2. Parametric bootstrap for fully parametric models.
    - Suppose $y|\mathbf{x} \sim F(\mathbf{x}, \boldsymbol{\theta}_0)$ and generate y_i^* by draws from $F(\mathbf{x}_i, \widehat{\boldsymbol{\theta}})$
  - 3. Residual bootstrap for regression with additive errors
    - Resample fitted residuals $\widehat{u}_1, ..., \widehat{u}_N$ to get $(\widehat{u}_1^*, ..., \widehat{u}_N^*)$ and form new $(y_1^*, \mathbf{x}_1), ..., (y_N^*, \mathbf{x}_N)$.
- Need to resample over i.i.d. observations
  - resample over clusters if data are clustered
    - But be careful if model includes cluster-specific fixed effects.
  - resample over moving blocks if data are serially correlated.

Bootstrap failure

- The bootstrap always provides estimates even when itmakes no sense
  - e.g. can always get bootstrap standard errors for the mean of a Cauchy sample, even though the mean of the Cuachy does not exist.
- The following are cases where standard bootstraps fail
  - so need to adjust standard bootstraps.
- GMM (and empirical likelihood) in over-identified models
  - For overidentified models need to recenter or use empirical likelihood.
- Nonparametric Regression:
  - Nonparametric density and regression estimators converge at rate less than root-N and are asymptotically biased.
  - This complicates inference such as confidence intervals.
- Non-Smooth Estimators: e.g. LAD.

Jackknife

- The jackknife uses a leave-one-out resampling scheme.
- The jackknife estimate of the variance of an estimator $\widehat{\theta}$ is

$$\widehat{\mathrm{V}}[\widehat{\theta}] = \tfrac{N-1}{N} \textstyle\sum_{i=1}^{N} (\widehat{\theta}_{(-i)} - \overline{\widehat{\theta}})^2, \quad \text{where } N^{-1} \textstyle\sum_i \widehat{\boldsymbol{\theta}}_{(-i)}.$$

  - where $\widehat{\theta}_{(-i)}$ is $\widehat{\theta}$ obtained from the sample with observation i omitted.
- The jackknife is a "rough and ready" method for bias reduction in many situations, but not the ideal method in any.
  - it can be viewed as a linear approximation of the bootstrap (Efron and Tibsharani (1993, p.146)).
  - it requires less computation than the bootstrap in small samples, as then $N < B$ is likely
  - but it is outperformed by the bootstrap as $B \rightarrow \infty$.
- E.g. `poisson docvis chronic, vce(jackknife)`

4. Bootstrap with asymptotic refinement

- The simplest bootstraps are no better than usual asymptotic theory
  - advantage is easy to implement, e.g. standard errors.
- More complicated bootstraps provide asymptotic refinement
  - this may provide a better finite-sample approximation.
- Conventional asymptotic tests (such as Wald test).
  - $\alpha =$ nominal size for a test, e.g. $\alpha = 0.05$.
  - Actual size$= \alpha + O(N^{-1/2})$.
- Tests with asymptotic refinement
  - Actual size$= \alpha + O(N^{-1})$.
  - asymptotic bias of size $O(N^{-1}) < O(N^{-1/2})$ is smaller asymptotically.
  - But need simulation studies to confirm finite sample gains.
    - ★ e.g. if $N = 100$ then $100/N = O(N^{-1}) > 5/\sqrt{N} = O(N^{-1/2})$.

Asymptotically pivotal statistic and studentized t-statistic

- Econometricians rarely use asymptotic refinement.
- Asymptotic refinement bootstraps an asymptotically pivotal statistic
  - this means limit distribution does not depend on unknown parameters.
- An estimator $\widehat{\theta} - \theta_0 \overset{a}{\sim} \mathcal{N}[0, \sigma^2_{\widehat{\theta}}]$ is not asymptotically pivotal
  - since $\sigma^2_{\widehat{\theta}}$ is an unknown parameter.
- But the studentized $t-$statistic is asymptotically pivotal
  - since $t = (\widehat{\theta} - \theta_0)/s_{\widehat{\theta}} \overset{a}{\sim} \mathcal{N}[0, 1]$ has no unknown parameters.
- So bootstrap Wald test statistic to get tests and confidence intervals with asymptotic refinement.
- Formally this is an empirical way of impementing an Edgewoth expansion
  - a higher order expansion than the usual one used for asymptotic theory
  - analogous to going out extra terms in a Taylor series expansion.

Percentile-t bootstrap

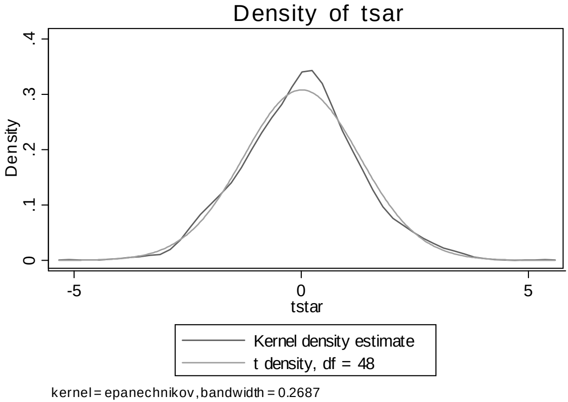
- Bootstrap $t = (\widehat{\theta} - \theta_0)/s_{\widehat{\theta}} \overset{a}{\sim} \mathcal{N}[0,1]$
  - by recomputing $t_b^* = (\widehat{\theta}_b - \widehat{\theta})/s_{\widehat{\theta}_b}$ where $\widehat{\theta}$ = original sample estimate.

```
. * Percentile-t for a single coefficient: Bootstrap the t statistic
. use bootdata.dta, clear

. quietly poisson docvis chronic, vce(robust)

. local theta = _b[chronic]

. local setheta = _se[chronic]

. bootstrap tstar=((_b[chronic]-`theta')/_se[chronic]), seed(10101)        ///
>   reps(999) nodots saving(percentilet, replace): poisson docvis chronic, ///
>   vce(robust)
(note: file percentilet.dta not found)

Bootstrap results                               Number of obs     =         50
                                                Replications      =        999

      command:  poisson docvis chronic, vce(robust)
        tstar:  (_b[chronic]-.9833014421442415)/_se[chronic]
```

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
tstar	0	1.288018	0.00	1.000	-2.52447	2.52447

Percentile-t bootstrap (continued)

- The 999 values of tstar $(= t_b^* = (\widehat{\theta}_b - \widehat{\theta})/s_{\widehat{\theta}_b})$ trace the bootstrap estimated density of the t-statistic .
- The plot is of the kernel density estimate and $T(48)$



Percentile-t bootstrap (continued)

- The test and confidence interval critical t-values are the 2.5 and 97.5 percentiles of t^*
- For a symmetric two-sided test of $H_0 : \theta = 0$ find the proportion of times that $|t^*| > |t|$

```
. * Percentile-t critical values, 95% confidence interval
. * and p-value for symmetric two-sided Wald test of H0: theta = 0
. use percentilet, clear
(bootstrap: poisson)

. centile  tstar, c(2.5, 97.5)

                                                      -- Binom. Interp. --
    Variable |       Obs  Percentile    Centile        [95% Conf. Interval]
-------------+-------------------------------------------------------------
       tstar |       999        2.5   -2.414266       -2.566386   -2.296212
             |                 97.5    2.661827        2.457073    3.011566

. display "95% conf. int. = (" .983-r(c_1)*0.515 ", " .983+r(c_2)*0.515 ")"
95% conf. int. = (2.226347, 2.3538408)

. * Percentile-t p-value for symmetric two-sided Wald test of H0: theta = 0
. quietly count if abs(tstatistic) < abs(tstar)

. display "tstatistic = " tstatistic "  and p-value = " r(N)/(_N+1)
tstatistic = 1.9075106  and p-value = .146
```

BC and BCa confidence interval

- The BC (bias-corrected) and BCa methods also provide asymptotic refinement.
- (N) is observed coefficient $\pm$ 1.96 $\times$ bootstrap s.e.
- (P) is 2.5 to 97.5 percentile of the bootstrap estimates $\widehat{\beta}_1^*, \ldots, \widehat{\beta}_B^*$.
- (BC) and (BCa) have asymptotic refinement.

```
. * Bootstrap confidence intervals: normal-based, percentile, BC, and BCa
. quietly poisson docvis chronic, vce(boot, reps(999) seed(10101) bca)

. estat bootstrap, all

Poisson regression                              Number of obs      =        50
                                                Replications       =       999

------------------------------------------------------------------------------
             |    Observed               Bootstrap
      docvis |       Coef.       Bias    Std. Err.  [95% Conf. Interval]
-------------+----------------------------------------------------------------
     chronic |   .98330144   .0132307   .54137854    -.077781   2.044384   (N)
             |                                     -.0139438   2.061742   (P)
             |                                      -.079079   2.019438  (BC)
             |                                      .0295944    2.08349 (BCa)
       _cons |   1.0316016  -.0769223   .35685342    .3321817   1.731021   (N)
             |                                       .186586   1.582409   (P)
             |                                       .268264   1.641356  (BC)
             |                                       .386773   1.771351 (BCa)
------------------------------------------------------------------------------
(N)    normal confidence interval
(P)    percentile confidence interval
(BC)   bias-corrected confidence interval
(BCa)  bias-corrected and accelerated confidence interval
```

5. Wild bootstrap

- The wild bootstrap is specific to linear regression.
- It conditions on the sample value of the $\mathbf{x}'s$.
- Only y is resampled. $\mathbf{x}$ is not resampled.
- In original sample do OLS of y_i on $\mathbf{x}_i$ and get residual $\widehat{u}_i = y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}$.
- In the b^{th} resample
  - set $y_{i,b} = \begin{cases} \mathbf{x}_i'\widehat{\boldsymbol{\beta}} + \widehat{u}_i & \text{with probability } 0.5 \\ \mathbf{x}_i'\widehat{\boldsymbol{\beta}} - \widehat{u}_i & \text{with probability } 0.5 \end{cases}$
  - do OLS regression using sample $(y_{1,b}, \mathbf{x}_1), \ldots., (y_{N,b}, \mathbf{x}_N)$ gives $t_b^* = (\widehat{\beta}_b - \widehat{\beta})/s_{\widehat{\beta}_b}$.
  - seems "wild" as $y_{i,b}$ can only take one of two values
  - but we are doing this for N observation so possibly as many as 2^N distinct samples.
- Gives an asymptotic refinement for OLS with heteroskedastic errors.

Wild bootstrap (continued)

- Not used much in practice for indpendent observations.
  - usually if N is so low that we need this bootstrap as asymptotic theory has kicked in then things are statistically insignificant.
- But with clustered data things can be highly statistically significant
  - yet tests have poor size if number of clusters G is small
  - Cameron, Gelbach and Miller proposed a cluster version of the Wild bootstrap where resample over clusters
  - i.e. $\mathbf{y}_{g,b} = \mathbf{X}_g'\widehat{\beta} + \widehat{\mathbf{u}}_g$ or $\mathbf{y}_{g,b} = \mathbf{X}_g'\widehat{\beta} - \widehat{\mathbf{u}}_g$ in cluster g
  - and then use the percentile-t method as before
- In theory one could instead use a simpler pairs cluster bootstrap
  - where resample clusters $(\mathbf{y}_g, \mathbf{X}_g)$ with replacement
  - but this worked poorly in Monte Carlos.

6. Stata commands

- Most commands have option `vce(bootstrap)` and `vce(jackknife)`
- For more complicated bootstraps write a program and use `bootstrap:`
- For replicability set the seed!!
- For published work the more bootstraps the better as the seed becomes less important.

7. References

- A. Colin Cameron and Pravin K. Trivedi (2005), *Microeconometrics: Methods and Applications (MMA)*, chapter 11, Cambridge Univ. Press.
- A. Colin Cameron and Pravin K. Trivedi (2009), *Microeconometrics using Stata (MUS)*, chapter 13, Stata Press.